

Guida breve alla digitalizzazione in biblioteca

di Piero Falchetta

I. Premessa

Sono innumerevoli, ormai, le esperienze e i progetti di acquisizione digitale di materiali – diversi per natura, caratteri intrinseci ed estrinseci, età, pregio e competenza – che le biblioteche di tutto il mondo hanno già da vari anni avviato, sia a scopo conservativo – per proteggere cioè gli originali dai rischi inevitabilmente connessi alla consultazione diretta – sia, e soprattutto, per consentire più facile e rapido accesso ai documenti da parte di un pubblico di lettori sempre più allargato.

Tali esperienze e progetti appaiono quanto mai diversificati, tanto in rapporto ai contenuti quanto al loro grado di fungibilità rispetto agli originali dai quali derivano, ma diversi sono soprattutto – sotto l'apparente similarità degli esiti funzionali – le tecniche e gli strumenti impiegati per creare/acquisire i dati in formato digitale, identificarli e descriverli sulla base di un sistema più o meno coordinato di metadati, archivarli in un modo sicuro e rispondente a principi di economicità d'accesso, distribuirli nelle diverse e continuamente rinnovate modalità offerte dall'evoluzione tecnologica.

Una simile disomogeneità si è finora determinata, in molti casi, per effetto di condizioni operative dettate innanzitutto, per così dire, *dall'esterno*, dove per esterno deve intendersi il processo di sviluppo delle tecnologie informatiche, processo che nella sua interazione con il mercato appare connotato da un livello di complessità tale da sottrarsi ininterrottamente ai tentativi di omogeneizzazione – se non di standardizzazione – richiesti o magari anche imposti dal di

fuori. In altre parole, le istituzioni culturali in generale e le biblioteche in particolare – per restare nell'ambito che più ci interessa – vedono mutare di continuo le regole del gioco e a tali cambiamenti e novità, per gli evidenti vantaggi tecnologici ed economici che questi spesso comportano, cercano di allineare le attività che siano ancora in fase di progettazione o di avvio, venendo in tal modo a contribuire all'accrescimento della diversificazione.

I rischi di obsolescenza dei dati e dei sistemi che li rendono operabili, e che tali rapidi cambiamenti lasciano facilmente prevedere, sono ormai da tempo oggetto di una discussione allargata; ma soltanto negli anni più recenti sono stati pubblicati articoli e studi contenenti proposte di coordinamento tecnico e organizzativo utili se non a scongiurare quanto meno a limitare gli effetti di obsolescenza che potranno essere indotti dagli sviluppi tecnologici futuri. Fra questi vanno senz'altro citate le *Guidelines for digitizing archival materials for electronic access* compilate in seno alla National Archives and Records Administration statunitense,¹ le direttive elaborate dalla Library of Congress per le proprie attività e i propri programmi di digitalizzazione,² le direttive, per venire all'Italia, contenute nella deliberazione AIPA 24/1998,³ il recentissimo *Studio sulla digitalizzazione* prodotto dalla Commissione per la conservazione del patrimonio librario nazionale istituita presso l'Ufficio centrale per i beni librari del Ministero per i beni e le attività culturali⁴ e la *Normativa per l'acquisizione digitale delle immagini fotografiche* dell'Istituto centrale per il catalogo e la documentazione dello stesso ministero.⁵

¹ Cfr. <<http://www.nara.gov/nara/vision/eap/eapspec.html>>.

² Tali direttive sono compendiate in CAROLINE R. ARMS, *Keeping memory alive. practices for preserving digital contents at the National Digital Library Program of the Library of Congress*, in "RLG DigiNews", 4 (2000), 3; questa rivista vive soltanto in formato elettronico e l'articolo citato può essere letto alla pagina <<http://www.rlg.org/preserv/diginews/diginews4-3.html>>.

³ Il testo è consultabile all'indirizzo <http://www.aipa.it/servizi3/normativa4/circolari2/db24_98.asp>; la deliberazione è ulteriormente commentata dalla nota esplicativa leggibile all'indirizzo <http://www.aipa.it/attivita2/standard5/archivia_zioneottica1/noteesplicative.asp>.

⁴ Cfr. la lettera circolare di detto ministero, n. 8012, del 4 luglio 2000.

⁵ Il documento, in formato PDF, è scaricabile dalla pagina <<http://www.iccd.beniculturali.it/standard/IMMAGINIDIGITALI.html>>.

Questi e altri consimili documenti prendono in considerazione soprattutto due aspetti: per prima cosa la definizione e la struttura dei metadati necessari non soltanto per l'identificazione intrinseca e amministrativa del documento digitale ma anche per la sua futura "riconoscibilità" da parte di programmi e sistemi che saranno con ogni probabilità diversi da quelli odierni; in secondo luogo la definizione di standard tecnici – quali ad esempio quelli relativi al formato delle immagini digitali – quanto più possibile trasparenti e omogenei.

Il carattere eminentemente pratico che il presente contributo intende avere non consente di approfondire più di tanto la questione relativa ai metadati, e ci limiteremo perciò a un paio di rapide osservazioni su questo argomento. La definizione di un set di metadati appare ancora lontana dal possedere l'autorità di uno standard, e tali prescrizioni appaiono del resto marginali nel contesto generale, in incessante rinnovamento, delle realizzazioni avviate e portate a compimento da istituti bibliotecari pubblici e privati, i quali operano sovente sulla scorta di risorse locali (economiche, tecnologiche e organizzative) che non sono in grado e/o non ritengono utile "allinearsi" entro sistemi normativi che possono comportare un notevole appesantimento delle operazioni, con conseguenti aumenti di spese e maggior grado di complessità delle piattaforme tecnologiche e della struttura organizzativa.

Basti per esempio tornare al caso della già citata deliberazione AIPA, formulata per favorire l'integrazione di regole tecniche e procedure allo scopo di "facilitare la diffusione dei sistemi di archiviazione ottica sia nel settore pubblico che nel settore privato" e di "determinare un coordinato sviluppo di tale sistema... onde ricondurre il tutto in una logica di comportamenti univoci che offrano affidabilità, definendo principi comportamentali uniformi nell'impiego della nuova tecnologia". L'applicazione puntuale della griglia normativa definita con tale testo appare tuttavia quanto mai onerosa, soprattutto nel caso in cui venisse sussunta da istituti e organismi bibliotecari di piccole dimensioni o dall'indirizzo marcatamente specialistico, con la conseguente prospettiva di un possibile detrimento dell'agilità delle procedure di archiviazione e accesso tipica del documento digitale.

Per quel che riguarda invece la definizione di indicazioni e parametri tecnici, la questione appare più semplice, in quanto il tentativo è di mettere a fuoco poche ed essenziali caratteristiche che il documento digitale elaborato oggi dovrebbe possedere per non perdere domani – a causa dell'obsolescenza di programmi e sistemi – la possibilità di essere "letto". Fanno parte di tali istruzioni i dati relativi alla qualità e al formato dei file grafici, alla regolazione del colore, alle tecnologie di acquisizione, ecc. Il rispetto di tali principi è del resto possibile in una misura abbastanza indipendente dalle scelte tecniche operate nel corso dei processi di digitalizzazione, le quali rimangono perciò a carico dei singoli programmi di attività. Queste dipendono innanzitutto dagli scopi che si intendono perseguire e inoltre dalla qualità e dalla quantità delle risorse economiche, tecniche e organizzative delle quali la singola istituzione dispone nella realizzazione di un determinato programma. Tanto più variegato e instabile appare il panorama delle

possibilità tecniche disponibili – le quali influenzano grandemente il numero e la consistenza delle procedure – tanto più importante sarà perciò la fase di progettazione delle attività di digitalizzazione, giacché essa non dovrà limitarsi a considerare gli scopi immediati di un singolo progetto, a valutare i benefici raggiungibili in rapporto alla quantità di risorse messe in campo, a mettere in conto tutte le attività ordinarie di manutenzione e aggiornamento necessarie dopo l'avvio del progetto, ma dovrà anche, per quanto possibile, creare condizioni per le quali le attività risultanti dal progetto possano essere in ogni momento riprese, modificate, aggiornate, trasferite senza che ciò comporti una revisione generale – costosa e difficile – dell'intero sistema di dati, di hardware e di software che costituisce e sostiene l'attività stessa.

Sullo sfondo di un simile scenario tecnologico e normativo, la presente *Guida breve* intende offrire un primo orientamento a quanti abbiano in programma attività di digitalizzazione di fondi documentali e librari conservati nelle biblioteche italiane. Tutti gli aspetti e le questioni fin qui accennati sono trattati nella *Guida* sotto tre profili: le risorse economiche, quelle tecnologiche e quelle organizzative, al fine di consentire che l'elaborazione di qualsiasi progetto contemperi al meglio le possibilità, gli scopi e le aspettative che i singoli istituti hanno e si prefiggono all'avvio dei programmi di digitalizzazione.

L'argomento è quanto mai complesso e la stessa rapidità dell'evoluzione tecnologica impedisce – come si è più volte ricordato – affermazioni ultime e definitive in tal senso; la guida va perciò intesa come una sorta di prontuario d'uso, o come una piattaforma progettuale sulla quale innestare le infinite variazioni necessarie e possibili. Nel tentativo di rendersi ancora più utile, la *Guida* riporta anche un questionario destinato a inventariare le risorse e le finalità di un determinato progetto, allo scopo di favorire la convergenza degli sforzi di quanti – dall'interno o dall'esterno dei vari istituti – collaboreranno alla realizzazione delle attività di digitalizzazione.

2. Programmare l'intervento

La programmazione di un'attività di digitalizzazione dalla portata più o meno ampia deve considerare innanzitutto gli scopi che si intendono perseguire e gli esiti attesi. Anche se in linea di principio tali attività rientrano in programmi generali di sfruttamento delle possibilità offerte dallo sviluppo tecnologico, gli scopi che un singolo istituto si propone di raggiungere possono essere quanto mai diversificati; si può infatti andare dalla creazione di copie d'archivio destinate *tout court* a preservare la memoria di determinati documenti al trasferimento in formato digitale di dati già conservati su altro supporto (ad es. il microfilm), dall'integrazione dei dati risultanti dall'attività su piattaforme digitali preesistenti all'intento primario di distribuire i dati digitali localmente o mediante accesso remoto (OPAC, Internet), dalla necessità di un aggiornamento tecnologico complessivo destinato a rinnovare la qualità dei servizi offerti e a costituire magari fonte di introiti, alla sperimentazione di metodi e tecniche propria di molti progetti-pilota. ➤

Altre questioni riguardano le previsioni circa l'ampiezza dell'eventuale utenza e le sue, per quanto prevedibili, aspettative; un conto è infatti offrire un servizio destinato soltanto all'accesso locale, un altro conto predisporre un sistema di distribuzione che garantisca all'utenza la continuità *world wide* del servizio. E ancora: i dati distribuiti avranno caratteristiche tali da poter essere in qualche misura sostitutivi dell'originale, oppure serviranno soltanto all'identificazione e al riconoscimento dei singoli documenti? E tali dati sono destinati a essere utilizzati soprattutto da un pubblico generico o si rivolgono invece a settori specializzati del mondo della ricerca? E l'accesso ai dati potrà essere consentito agli utenti senza mediazione alcuna oppure sarà necessario attivare un servizio assistito?

Le risposte a questi ed altri simili interrogativi, qualora essi vengano esplicitati nella fase preparatoria, serviranno a circoscrivere il campo di applicazione del progetto, a commisurarne meglio i risultati in rapporto alle risorse, ad accrescerne l'efficacia e in certa misura a evitare inutili appesantimenti delle procedure, che si traducono inevitabilmente in aggravio dei costi.

2.1 Ricognizione e individuazione dei fondi

Questa fase è senza dubbio una delle più importanti e delicate dell'intero processo. Sia che l'istituto privilegi l'aspetto della conservazione, sia che miri i suoi sforzi innanzitutto alla creazione di maggiori possibilità di accesso ai dati, molto dipenderà – in termini di risorse da impiegare – dalle caratteristiche fisiche dei documenti che saranno sottoposti alla digitalizzazione.

La ricognizione consisterà non soltanto nell'individuazione dei fondi che si vogliono e/o si possono trattare, ma anche nella quantificazione il più possibile precisa dei materiali in rapporto alle tipologie alle quali appartengono. È infatti chiaro come la digitalizzazione, ad esempio, di un numero *n* di pagine a stampa legate in volume inciderà diversamente su costi, organizzazione e attrezzature necessarie da quella di documenti ad alto contenuto visuale (grafica, cartografia, ecc.) o di antiche opere manoscritte che non presentino magari problema alcuno per quel che riguarda la leggibilità ma che dovranno essere maneggiate con cure particolari, imponendo così un notevole prolungamento delle operazioni di ripresa.

Può perciò apparire in alcuni casi preferibile, in termini di utilità ed efficacia del servizio finale, trattare documenti più "facili", giacché si ottiene con questi una più vantaggiosa economia di scala. Certo gli istituti culturali, e le biblioteche in particolare, non sono quasi mai, in Italia e in Europa, delle imprese sottoposte alla legge del profitto, e l'apparente gratuità di una parte del loro ciclo di attività è solitamente motivata da ragioni di servizio culturale. Si è tuttavia assistito più di qualche volta al dispiegamento di notevoli mezzi e risorse che si sono limitati a produrre esili risultati, generando l'impressione di un impiego non ottimale delle disponibilità. La tecnologia digitale è infatti adatta a essere impiegata per realizzazioni di grande mole, e gli investimenti iniziali, spesso ingenti, necessari per metterla in atto possono apparire non convenienti se il prodotto finale non è rilevante anche sotto il profilo della quantità.

2.2 Le risorse umane e organizzative

La valutazione delle risorse umane e organizzative delle quali si dispone è un punto della massima importanza. Sia che si ricorra a personale interno all'istituto, sia che ci si appoggi a servizi esterni, bisognerà prevedere i seguenti passaggi organizzativi, a ciascuno dei quali corrispondono competenze diverse:

a) nella fase di ricognizione e individuazione dei fondi, condotta con ogni evidenza dai responsabili interni, sarà importante valutare non soltanto le tipologie documentali da trattare, ma anche la dislocazione dei materiali; è infatti necessario che tali materiali siano trasportati dai magazzini dove sono abitualmente conservati fino al luogo (per lo più interno, ma in alcuni casi anche esterno all'istituto) dove avverranno le operazioni di digitalizzazione. La presa e la ricollocazione dei materiali richiederà senza dubbio personale da adibire allo scopo, in misura tanto più grande quanto maggiori saranno sia la mole complessiva dei lotti da trattare sia la fragilità e il pregio dei documenti da digitalizzare. Sarà perciò opportuno sentire anche il parere dei responsabili della conservazione e/o del restauro, i quali dovranno valutare le eventuali controindicazioni al trattamento di materiali che per intrinseca fragilità, misure molto ampie, presenza di piegature e/o lacerazioni, cattivo stato delle legature, conservazione in rotolo, presenza di dorature passibili di distacco, rischi connessi all'esposizione alla luce ecc. possano subire danni in seguito al trattamento;

b) i responsabili di biblioteca dovranno definire l'organizzazione dell'indicizzazione dei dati ed, eventualmente, la struttura dei metadati. L'identificazione dei record digitali dovrà essere infatti tale da consentire la rapida e precisa associazione degli stessi con quelli descrittivi e/o di catalogazione. I materiali da trattare dovranno perciò essere in qualche modo inventariati e descritti – se non catalogati – a cura di personale espressamente addestrato, e tali operazioni dovranno essere condotte precedentemente all'avvio delle operazioni di ripresa;

c) la fase della digitalizzazione vera e propria potrà essere condotta tanto da personale interno quanto esterno; sia che le riprese siano effettuate con tradizionale metodo fotografico sia con digitalizzazione diretta per mezzo di scanner o testata digitale, tali operazioni comportano il contributo di tecnici fotografici, tecnici HW-SW capaci di trarre il maggior vantaggio dalle strumentazioni disponibili e incaricati di effettuare le operazioni di backup e di riversamento dei dati sul supporto finale. Tali tecnici dovranno lavorare di concerto con il personale addetto alla presa e ricollocazione, in modo che le operazioni non subiscano gravose interruzioni. I dati dovranno inoltre essere revisionati a campione dai responsabili di biblioteca, in modo da verificare la presenza di eventuali errori e difetti tali da rendere necessarie nuove riprese.

2.3 Le risorse di gestione (post-produzione)

Dopo l'archiviazione dei dati sul supporto finale, questi dovranno essere trattati da tecnici incaricati di integrarli in un sistema preesistente o costituito *ad hoc*. Anche sotto l'aspetto della gestione dei dati, questo passaggio comporta notevoli

conseguenze per l'assetto generale del sistema. Diverso è infatti il caso in cui tali operazioni siano controllate e condotte o meno dal personale interno all'istituto, il quale sarà perciò in grado di seguire passo passo l'evoluzione del sistema, di intervenire con continuità per tutti i necessari riallineamenti, di mantenere nel sistema condizioni di non-incompatibilità sia verso l'esistente sia verso eventuali futuri sviluppi.

Altro caso (il più frequente) è invece quello nel quale le fasi di post-produzione e di manutenzione del sistema siano affidati a servizi esterni. Sarà della massima importanza, in tale evenienza, assicurarsi fin dal momento della stesura del contratto che la presenza dei tecnici incaricati della post-produzione e della manutenzione sia garantita con tempestività e continuità. Nel momento in cui i dati acquisiti saranno infatti offerti in forma di servizio all'utenza, risulteranno quanto mai dannosi per la qualità dello stesso e per l'immagine complessiva dell'istituto le spesso inevitabili interruzioni, i malfunzionamenti, gli incidenti di vario genere che tanto spesso affliggono i sistemi informatici complessi. Tali inconvenienti potranno essere in parte attenuati se i tecnici esterni avranno la possibilità di intervenire con rapidità, e ciò dipende in primo luogo dalla vicinanza geografica della ditta incaricata all'istituto. La costante cura che tali realizzazioni richiedono risulta infatti assai più gravosa nel caso in cui gli interventi debbano essere cadenzati nel tempo, o qualora il contratto preveda la manutenzione nella sola forma *on line*. Molto spesso il personale di biblioteca non possiede infatti le competenze necessarie per intervenire in profondità sui sistemi informatici, e il rischio di paralisi degli stessi aumenta, si potrebbe dire, in rapporto alla distanza geografica tra la ditta incaricata e la biblioteca. Appare perciò preferibile, a volte, affidare tali servizi a ditte di piccole dimensioni, che siano però presenti sul territorio, piuttosto che a grandi imprese le quali, se è vero che possono offrire maggiori garanzie per quel che riguarda l'aggiornamento tecnologico, la specializzazione del personale, la vasta conoscenza ed esperienza, si trovano spesso a disagio nella realizzazione di progetti di portata limitata, di produzioni di carattere "artigianale", quali sono per la maggior parte quelle realizzate nell'ambito delle biblioteche italiane non di prima grandezza.

Un altro punto da considerare – sebbene non di primaria importanza e tuttavia non influente – riguarda la distribuzione fisica dei dati. Possono allora darsi due casi: nel primo caso i dati sono costantemente *on line* e non è perciò necessaria alcuna operazione di dislocamento fisico dei supporti. Nel secondo caso, assai frequente, i dati sono archiviati su supporti *off line* (cd-rom e dvd non serviti da librerie, ecc.), e si dovrà perciò provvedere all'istruzione di personale incaricato di operazioni semplici, quali la presa e la ricollocazione dei supporti (la quale comporta tuttavia notevoli rischi per la conservazione degli stessi), l'attivazione di una postazione di consultazione e dei programmi destinati alla visualizzazione dei dati. Tali passaggi devono essere programmati con grande cura, soprattutto nel caso in cui il servizio sia stato ufficialmente istituito, venendo in tal modo a ingenerare curiosità e aspettative da parte dell'utenza.

Un'ultima considerazione riguarda ancora l'accesso ai dati. L'interoperabilità degli archivi può essere garantita da interfacce *user friendly*, come si dice, le quali non necessiteranno di assistenza se non in casi particolari, ovvero può rendersi necessaria una forma di assistenza da parte del personale di biblioteca. Tali compiti di assistenza dovranno allora riguardare non soltanto la parte tecnica (in sostanza, l'impiego dei programmi e delle attrezzature), ma anche i contenuti dell'archivio, in modo non molto diverso da quanto avviene per la normale assistenza bibliografica di sala. Se per quest'ultima parte non vi è bisogno di un particolare addestramento del personale addetto, per quel che riguarda l'assistenza tecnica gli addetti al servizio dovranno evidentemente essere portati a un'adeguata formazione di base, la quale dovrà essere poi continuamente aggiornata per effetto degli inevitabili cambiamenti che il sistema potrà adottare o richiedere.

3. Determinazione dei contenuti tecnici dell'intervento

I contenuti tecnici dell'intervento potranno essere determinati sulla base di alcune questioni sostanziali, ovvero in rapporto ai costi, allo scopo e alle risorse tecnologiche e di gestione disponibili. In questa fase della programmazione sarà inoltre necessaria una stretta collaborazione fra il personale responsabile degli indirizzi generali e delle scelte tecniche e quello preposto all'espletamento degli atti amministrativi. La realizzazione di un progetto prevede infatti non soltanto la stipula di contratti di vario genere e natura, il lancio di bandi, di gare o di richieste d'offerta, ma richiede anche attenzione alla trasparenza assoluta degli atti, poiché questi potrebbero essere impugnati – il caso non è infrequente – dai partecipanti esclusi; parimenti dovrà essere messa in conto anche la redazione di rendiconti e relazioni periodiche, richiesti come adempimenti obbligatori – molteplici sono le prescrizioni in tal senso – dalle amministrazioni locali o da quelle centrali.

3.1 Determinazione dei contenuti tecnici dell'intervento in rapporto ai costi

I costi di realizzazione di un progetto di digitalizzazione dipendono da molteplici fattori, i quali sono tuttavia riconducibili a due elementi fondamentali: l'ampiezza e la durata del progetto da un lato, l'evoluzione tecnologica e organizzativa dell'istituto dall'altro. Tali fattori sono interdipendenti e, in particolare, sono condizionati innanzitutto dalle scelte tecniche e operative, le quali hanno effetto non soltanto sul costo puro delle operazioni di ripresa e digitalizzazione, ma anche, e in misura notevole, sul costo da sostenersi per la gestione dei dati.

In assenza di stime sistematiche e statistiche di fonte italiana, non sarà allora inutile rifarsi per un raffronto di massima a un recente articolo americano,⁶ nel quale l'incidenza dei costi in alcuni progetti di digitalizzazione realizzati dalla Library ➤

⁶ STEVEN PUGLIA, *The costs of digital imaging projects*, "RLG DigiNews", 3 (1999), 5, consultabile alla pagina <<http://www.rlg.org/preserv/diginews/diginews3-5.html>>.

of Congress e dai National Archives è stata analizzata con grande dettaglio. Fra le molte stime presentate nell'articolo vi sono anche quelle relative al costo medio per immagine finale – un costo che appare compreso fra i 23 e i 29 dollari. Ciò che interessa tuttavia notare, al di là del puro dato quantitativo, è la ripartizione dei costi per le diverse operazioni effettuate, in base alla quale si constata quanto segue:

- un terzo circa del costo è destinato alle spese per le operazioni di digitalizzazione;

- un altro terzo circa è impiegato per la creazione dei metadati, notando però che in tale voce sono comprese anche le operazioni di indicizzazione e soprattutto quelle di catalogazione;

- l'ultimo terzo circa è infine assorbito dalle spese di gestione, incluse quelle amministrative e di controllo di qualità.

Quali indicazioni si possono trarre da tali cifre? Innanzitutto va notato come queste non comprendano le dotazioni HW e SW necessarie, mentre tale capitolo di spesa preliminare comporta un investimento che può coprire una notevole parte del finanziamento di progetto. Le spese per le attrezzature e i programmi sono tuttavia da considerarsi investimenti a lungo termine, poiché la piattaforma tecnologica acquisita potrà naturalmente essere impiegata per ulteriori realizzazioni; è in tal senso quanto mai opportuno valutare l'espansibilità futura dei sistemi di dati, e le possibilità che ogni istituto ritenga di avere sia in termini di contenuti, per i quali sia vantaggioso un passaggio al formato digitale, sia in termini di sviluppo tecnologico.

Un altro costo la cui incidenza può variare in misura notevole è quello relativo alla creazione dei metadati e, in particolare, alla catalogazione dei documenti. Se per questi ultimi esiste già una forma di catalogazione implementabile nel sistema con procedure relativamente semplici, i costi per la creazione dei metadati saranno drasticamente ridotti, mentre la loro incidenza sarà assai più alta nel caso contrario.

E ancora: a fronte di un progetto avviato grazie a un finanziamento straordinario, ovvero fuori dai capitoli della contabilità ordinaria dell'istituto – e questo è il caso che più frequentemente si dà nelle biblioteche italiane – andranno valutate in modo diverso le spese per la creazione dei metadati, e della catalogazione in particolare, a seconda del fatto che questa sia condotta a cura del personale interno o sia invece affidata a ditte esterne incaricate di trattare i lotti interessati. È chiaro che in questo secondo caso il finanziamento iniziale dovrà comprendere anche tali spese, che sono spesso ingenti a causa della grande mole di documenti che può essere trattata in un progetto di digitalizzazione di medie o grandi dimensioni, e dei tempi piuttosto lunghi che la catalogazione richiede.

In linea di massima, le spese complessive saranno comunque ripartite fra le seguenti voci:

- acquisto di attrezzature HW e SW per la creazione *ex novo* della piattaforma o per l'aggiornamento e/o il potenziamento di quella esistente;

- digitalizzazione;

- creazione di copia di sicurezza dei dati e di copia/e destinata/e alla distribuzione, con eventuale diversificazione del formato degli stessi;

- creazione dei metadati (indicizzazione, amministrativi e di catalogazione);

- funzionalizzazione del sistema, con particolare riferimento all'associazione dati digitali-metadati;

- manutenzione del sistema, compreso il backup periodico.

Un'ulteriore incidenza sui costi potrà inoltre essere determinata dal mancato rispetto della naturale consequenzialità delle procedure; se la creazione dei metadati sarà effettuata prima della digitalizzazione, si potranno evitare tutti i ritardi e gli aggravii comportati, nel caso contrario, dalla necessità di recuperare i dati necessari per la corretta associazione metadati-dati digitali e per il loro allineamento nel sistema.

3.2 Le finalità

Come si è in precedenza già accennato, le finalità generalmente perseguite con l'attuazione di progetti di digitalizzazione sono di due specie principali; da una parte vi sono quelle che intendono assicurare innanzitutto la conservazione della memoria, dall'altra quelle che finalizzano la riproduzione dei dati ottenuti alla loro distribuzione su scala più o meno ampia.

Nel primo caso la preoccupazione principale sarà quella di ottenere dati in un formato il più possibile "trasparente" ai futuri cambiamenti tecnologici – in altre parole, di archiviare i dati in formati non soltanto largamente diffusi, ma che riducano quanto più possibile le perdite di informazione legate ai diversi tipi e fattori di compressione, ovvero in formati universalmente validi in tal senso (relativamente alle immagini), quali i formati TIFF non compresso e PDF. Nel secondo caso, a tale assicurazione di fondo dovranno aggiungersi molte altre considerazioni, che sono indotte dall'opportunità di commisurare l'impiego delle risorse disponibili agli scopi realisticamente perseguibili dal progetto. I principali elementi discriminanti dal punto di vista tecnico e tecnologico sono i seguenti:

Ambito della distribuzione dei dati. Per una modalità di distribuzione locale del tipo *stand alone* non sarà necessario nessun sistema servente, ma soltanto una o più postazioni di accesso, magari dedicate; in questo caso la complessità del sistema sarà assai ridotta, in quanto le procedure potranno in sostanza ridursi alle fasi di acquisizione e archiviazione, mentre la distribuzione dei dati potrà avere carattere manuale o automatico, a seconda che il supporto di archiviazione viva *on line* o *off line*. Un'architettura di questo genere non necessita, al livello più basso, di programmi per la gestione dei metadati e l'associazione degli stessi con i dati digitali, ma può servirsi di semplici elenchi che descrivano i contenuti dell'archivio.

Nel caso in cui la distribuzione locale sia effettuata per mezzo di una rete, sono invece necessari l'appoggio a una stazione servente destinata a tale scopo e la presenza di programmi per la gestione delle comunicazioni e dei dati. Converrà in questo caso valutare la convenienza strategica di inserire il progetto in un'architettura preesistente, nel caso questa esista, o invece di crearne una che abbia un carattere dedicato, e che sia perciò dipendente dal server principale soltanto per il passaggio delle comunicazioni. Questa seconda possibilità può rivelarsi in alcuni casi vantaggiosa, in quanto consente l'adozione di sistemi informatici di minor complessità e costo, senza tuttavia dover rinunciare alla garanzia della sicurezza. Sottosistemi serventi dedicati, costruiti

ad esempio su piattaforme Linux o Windows NT Server, possono offrire a realtà di piccole e medie dimensioni molti vantaggi in termini di agilità ed economicità di gestione.

Nel caso infine in cui l'ambito della distribuzione preveda la possibilità di accesso remoto, potranno darsi le modalità di accesso OPAC e Internet, alternativamente o congiuntamente. La prima richiede architetture HW-SW fortemente strutturate e altamente complesse, il cui sviluppo è in buona parte regolamentato da norme e indicazioni emesse da istituti preposti al coordinamento e al controllo, come l'ICCU. La finalità in questo caso è principalmente, in Italia, la distribuzione dei dati all'interno di SBN per mezzo di protocolli di comunicazione ad alto grado di standardizzazione e di controllo, come ad esempio il protocollo Z3950. La scelta o l'esigenza di inserire un progetto di digitalizzazione in un ambiente già così notevolmente strutturato e normalizzato ridurrà la questione relativa ai contenuti tecnici al rispetto delle procedure di integrazione dei dati nel sistema esistente.

Molto più aperta la seconda soluzione, ovvero Internet, la quale non offre per contro le garanzie di omogeneità e la standardizzazione assicurate dalla prima. La stazione servente dovrà in questo caso possedere funzioni di web server del tutto automatizzate, in modo da garantire la regolare e continua distribuzione dei dati. Resta a carico di questa soluzione l'interfacciamento delle modalità di accesso con il database contenente i dati; su questo versante le soluzioni tecnologiche sono quanto mai diversificate, e ciò appare un vantaggio soprattutto nel caso di realtà e progetti dalle ridotte dimensioni, nelle quali simili sistemi possono essere configurati con un alto grado di specificità e precisione.

Funzione dei dati. Come si è in precedenza accennato, le caratteristiche tecniche dei dati digitali potranno diversificarsi a seconda del loro presumibile impiego. Con l'avvertenza che questa *Guida* esamina soprattutto – giacché questo è il caso di gran lunga più frequente nell'ambito delle biblioteche – l'acquisizione di dati in formato immagine, e non tratta perciò le acquisizioni OCR o la distribuzione dei file audio o altro, è possibile abbozzare una suddivisione per tipi, in funzione generale del loro utilizzo.

Due sono le condizioni d'accesso principali che possono darsi, separatamente o congiuntamente, in un qualsiasi sistema di distribuzione di dati digitali; la prima si limiterà a consentire la ricerca e il riconoscimento delle immagini, la seconda offrirà una riproduzione di qualità tale da poter essere proposta in sostituzione dell'originale, oltre a permettere una gestione delle immagini stesse, e darà inoltre modo di governare funzioni di stampa evolute.

Nel primo caso, utile per le funzioni di ricerca veloce, il formato quasi universalmente impiegato è il JPEG o JPG, che presenta i seguenti vantaggi: file di dimensioni piuttosto contenute e di conseguenza drastica riduzione della mole degli archivi e possibilità reale di accesso remoto, alta definizione del colore, adattabilità a piattaforme diverse e riconoscibilità pressoché universale. A fronte di tali vantaggi vi è però la perdita più o meno grande di informazioni, in quanto questo formato lavora per algoritmi di compressione che distruggono una notevole quantità di informazioni apparentemente inessenziali. A causa di ciò, le immagini JPG non potranno essere impiegate a scopo di archiviazione di qualità, né più di tanto ingrandite per un esame dettagliato,

e non potranno perciò essere offerte, nella gran parte dei casi, quali sostitutive dell'originale. Gli archivi JPG si prestano pertanto bene alle funzioni necessarie per le semplici operazioni di associazione fra descrizione e immagine, alla creazione di cataloghi visuali, alla riproduzione di materiale a stampa di agevole lettura, per il quale la distribuzione dei contenuti prevalga di gran lunga sull'informazione circa un esemplare determinato di un certo documento inteso quale insieme indivisibile di caratteri estrinseci e intrinseci.

D'altro canto, la ridotta quantità di memoria impegnata dai file JPG ne facilita grandemente l'archiviazione e la distribuzione. Potendo predeterminare il grado di compressione adottato, sarà infatti possibile trasformare un'immagine non compressa di molte decine di MB in una copia JPG assai "leggera", che occuperà soltanto poche centinaia di Kb di memoria. A semplice titolo d'esempio, se si avranno immagini non compresse del "peso" di 30 MB ciascuna, sarà possibile archivarle in un cd-rom non più d'una ventina, mentre sullo stesso supporto troveranno posto centinaia di immagini nel formato JPG. Archivi costituiti da migliaia di immagini JPG potranno essere così ospitati senza difficoltà in un qualsiasi hard disk, con gli evidenti vantaggi che ciò comporta in termini di agilità e di semplicità di gestione, nonché di velocità di accesso. Le strumentazioni HW necessarie in tal caso potranno così essere assai più ridotte, e il loro costo inciderà in misura molto minore sull'insieme del finanziamento.

Qualora le immagini digitali debbano invece possedere caratteristiche tali da consentirne una piena lettura, il formato prescelto dovrà senz'altro assicurare che non vi sia perdita di informazioni dovuta a procedure di compressione. Tali garanzie sono state individuate dall'uso corrente e dalle indicazioni provenienti da fonti diverse nell'impiego del formato TIFF, che sarà senz'altro da adottare per la riproduzione ad alta risoluzione delle immagini. Sarà in tal modo possibile proporre la copia digitale in luogo del documento originale – se non in tutti, quanto meno in gran parte dei casi.

L'adozione del formato TIFF richiede tuttavia l'impiego di notevolissime quantità di memoria, e l'archiviazione dei dati dovrà essere perciò programmata con cura, giacché gli archivi potranno facilmente raggiungere la misura delle centinaia di GB o addirittura delle decine di TB. Anche se la tecnologia sta compiendo in questo campo passi da gigante – grazie alla disponibilità di hard disk di grande capacità uniti a cascata, di librerie che possono alloggiare centinaia di cd-rom o di dvd e, prossimamente, di quella che si preannuncia come una novità di grandissimo interesse, ovvero la tecnologia FMD – si tratterà pur sempre di archiviare e distribuire in modo sicuro ed efficiente i dati relativi a migliaia o decine di migliaia di documenti grafici, ciascuno dei quali potrà "pesare" da poche unità fino ad alcune centinaia di MB. I supporti finali, qualunque essi siano, tenderanno così a moltiplicarsi, e a porre in tal modo non pochi problemi per la loro gestione organizzativa e tecnologica.

Le possibilità tecniche che potranno essere individuate dovranno perciò non soltanto garantire l'espansibilità degli archivi ma anche garantire la trasportabilità dei dati che, si ricordi, nel caso del formato TIFF costituiscono, in rapporto all'originale, le copie a maggior tasso di contenuti in- ➤

trinseci ed estrinseci. Il gruppo di studio a tale scopo costituito presso la Library of Congress ha individuato⁷ cinque metodi per la conservazione il più possibile prolungata dei dati nel tempo, e anche se tali indicazioni sono state formulate per la salvaguardia dei dati digitali in generale, a ben maggior ragione esse saranno valide nel caso di dati "di pregio", quali sono appunto quelli che duplicano i documenti originali nel formato TIFF.

1) *Supporti*: principio guida sarà la longevità e l'indipendenza dei supporti di archiviazione dall'evoluzione tecnologica. La longevità è assicurata, oltre che dall'impiego di materiali di qualità, dall'accuratezza delle operazioni di archiviazione, mentre l'indipendenza dalla tecnologia sarà tanto più alta quanto più ci si affiderà a formati "trasparenti", dei quali si conoscano le specifiche e per i quali vi sia un adeguato corredo di metadati.

2) *Refresh*: il refresh periodico dei dati (che comprende qualsiasi operazione di copia degli stessi) offre buone garanzie contro il decadimento fisico dei supporti e quello, per lo più accidentale, dei dati stessi.

3) *Aggiornamento*: la conversione dei dati esistenti in altro formato, tecnologicamente più avanzato, moltiplica le possibilità di sopravvivenza dei dati.

4) *Emulazione*: questo metodo consiste nel simulare le potenzialità di una tecnologia precedente all'interno di un sistema che appartenga a una generazione tecnologica più aggiornata; si verificherà così la "tenuta" dei dati nel corso di tali passaggi.

5) *Archeologia digitale*: ultima ratio, la ricerca di soluzioni per il recupero di tutti quei dati per i quali non si sia provveduto a procedure di archiviazione sistematiche e "trasparenti".

3.3 Indicazioni tecniche

I contenuti tecnologici di un sistema completo per l'acquisizione, l'archiviazione e la distribuzione di dati digitali sono costituiti da un insieme di attrezzature HW e di programmi, necessari per il raggiungimento delle finalità previste dal progetto. La determinazione dei contenuti tecnici dell'intervento dipende, come si è visto, da molteplici fattori di ordine strutturale e metodologico, ma le scelte operate dovranno comunque portare a una definizione tecnica del sistema tale da garantire i migliori risultati. Nel corso delle varie fasi dell'intervento sarà perciò necessario far ricorso ad alcune delle tecnologie di seguito indicate:

a) *Acquisizione fotografica*. Nel caso in cui l'acquisizione preveda anche il passaggio attraverso la ripresa fotografica tradizionale, potranno distinguersi tre tecniche principali:

- ripresa su pellicola 35 mm: può risultare molto efficace ed economica nel caso di materiali di facile leggibilità e di limitato valore estrinseco, come ad esempio i periodici, i repertori di uso frequente, e tutti quei documenti per i quali sia opportuna una più ampia disponibilità e facilità d'accesso senza che ciò comporti eccessivi rischi alla loro conservazione – trattandosi spesso, in questi casi, di materiali scarsamente resistenti alla frequente consultazione;
- ripresa su pellicola 6 x 6: indicata per tutti quei materiali

di pregio, manoscritti e a stampa, le cui dimensioni non superino i 60/70 cm sul lato più lungo, e la cui leggibilità non sia gravemente compromessa da scoloriture degli inchiostri, scurimenti accentuati della carta, dimensioni particolarmente ridotte delle scritture;

– ripresa su pellicola 10 x 12: indicata per materiali le cui dimensioni superino i 60/70 cm sul lato più lungo, e in tutti quei casi in cui la leggibilità del documento risulti ardua o gravemente compromessa.

Il passaggio dall'uno all'altro di questi supporti comporta aumenti geometrici dei costi, sia a causa del maggior valore dei materiali fotografici impiegati (fotocamere e pellicole), sia per il dilatarsi dei tempi di ripresa, sia infine per il maggior grado di specializzazione richiesto da parte del personale addetto a tali operazioni.

b) *Acquisizione digitale*. L'acquisizione digitale dei documenti potrà essere di due specie, ovvero indiretta e diretta:

– acquisizione indiretta: è quella ottenuta a partire dalle riprese fotografiche precedentemente eseguite. Nel caso dei film a 35 mm tale operazione potrà avvalersi dell'impiego di speciali scanner capaci di acquisire in modo pressoché automatico le immagini impresse sulla pellicola. Tali attrezzature sono in grado di lavorare ad alta velocità, permettendo la rapida costituzione di archivi digitali di notevoli dimensioni. Nel caso invece di riprese su pellicola 6 x 6 o 10 x 12, l'acquisizione necessiterà di scanner piani o a tamburo; l'impiego di tali attrezzature comporta una incidenza molto maggiore dei costi a causa del forte rallentamento delle operazioni, dovuto sia al fatto che le procedure saranno per lo più manuali, sia all'allungamento dei tempi richiesti per la scansione, i quali potranno passare dalle decine di secondi/immagine (film 35 mm) alle decine di minuti/immagine (pellicola 10 x 12).

Va inoltre ricordato quanto segue: la leggibilità di un documento digitale è determinata dal suo livello di risoluzione, ovvero, nel caso dell'acquisizione indiretta, dal rapporto definito tra dimensioni dell'originale, dimensioni del supporto fotografico e risoluzione di acquisizione. Per maggior chiarezza si può portare il seguente esempio: se l'immagine digitale di un documento che misuri 50 cm sul lato più lungo e che sia stato fotografato su pellicola 6 x 6 (con area utile per la scansione uguale a 5 x 5 cm circa) avrà la dimensione di 3.000 punti sul lato più lungo, ciò significa che la sua risoluzione finale sarà di 600 punti per centimetro, ovvero di 1.524 dpi (punti per pollice) rispetto alla pellicola, ma di soli 152 dpi (corrispondenti a circa 60 punti per cm) rispetto all'originale, e tale risoluzione non permetterà in molti casi una lettura soddisfacente del documento. Bisognerà insomma tener conto delle dimensioni del documento originale, per poter calibrare al meglio la risoluzione di acquisizione; la pellicola fotografica ha infatti dimensioni sempre uguali, ma il numero di informazioni che dovranno essere in essa contenute varierà grandemente col variare delle misure dell'originale; quanto più grande sarà, tanto più le informazioni saranno "concentrate" sulla pellicola, e la loro completa restituzione digitale sarà perciò possibile soltanto innalzando il livello della risoluzione di acquisizione;

⁷ Si veda il già citato articolo *Keeping Memory Alive*.

– acquisizione diretta: questo metodo comporta una sensibilissima riduzione dei costi operativi rispetto a quello precedente, in quanto elimina il laborioso e oneroso passaggio della ripresa fotografica. D'altro canto, le attrezzature necessarie per realizzarlo sono in genere assai costose, specie quando si tratti di apparecchiature capaci di riprendere originali di grande formato (A1 o superiore). Il loro acquisto risulterà perciò conveniente in quei casi in cui se ne preveda un impiego prolungato nel tempo e in progetti di digitalizzazione di estesi fondi librari e documentari.

Le tecnologie di acquisizione diretta sono in sostanza due: la prima si avvale dell'impiego di testate digitali montate su fotocamere o dedicate, la seconda opera la scansione per mezzo del passaggio sulla superficie del documento di un fotosensore a led – non molto diversamente, dal punto di vista meccanico, dal modo in cui lavora una comune fotocopiatrice.

Le differenze più importanti fra i due metodi sono le seguenti: la testata montata su fotocamera consente l'acquisizione di documenti di dimensioni molto grandi (oltre il formato A0); la testata dedicata è in genere montata su banchi ottici meccanizzati, che permettono di lavorare con buone garanzie di sicurezza anche su materiali delicati, quali possono essere i volumi di notevole spessore o la cui legatura risulti fragile; l'illuminazione necessaria per la ripresa può essere contenuta entro livelli di sicurezza; l'acquisizione è molto rapida, potendo in alcuni casi misurarsi in poche decine di secondi, anche nel caso di documenti di grande formato. Per contro, il metodo di scansione con fotosensore a led consente di ottenere, in termini di qualità, risultati del tutto paragonabili a quelli del sistema precedente, e in alcuni casi anche con un certo risparmio di spesa. Vi sarà tuttavia un limite nelle dimensioni degli originali che possono essere acquisiti e, soprattutto, vi saranno minori garanzie per quel che riguarda la tutela degli originali. I fotosensori a led producono infatti una notevole quantità di calore che viene irradiato a pochi centimetri di distanza dalla superficie dell'originale, condizioni che non si rivelano certo adatte nel caso di documenti antichi e di pregio, quali pergamene, miniature, disegni, o la cui carta sia fragile e gli inchiostri decolorati. Tali apparecchiature non risultano inoltre particolarmente indicate nel caso di digitalizzazione da volumi di grosso spessore, mentre si prestano assai meglio alla ripresa da documenti a fogli sciolti. Sarà quindi compito di ogni istituto valutare pregi e difetti di tali metodi anche in relazione alle tipologie e alle caratteristiche dei materiali che dovranno essere acquisiti in digitale.

Quanto poi ai livelli di risoluzione ottica sufficienti a garantire una buona leggibilità del documento digitale, si può tener conto di alcune indicazioni di massima (le caratteristiche e lo stato di conservazione del documento originale possono infatti richiedere variazioni anche sensibili rispetto ai valori di seguito indicati).

I documenti il cui valore informativo sia determinato non soltanto dai contenuti dell'informazione, ma anche dalla for-

ma intrinseca ed estrinseca nella quale questa si presenta – e in questa suddivisione possono rientrare tutti i documenti antichi e di pregio, sia manoscritti sia a stampa, e inoltre la grafica antica e moderna e quella particolare sintesi di testo e immagine che costituisce la caratteristica principale della cartografia – dovranno essere ripresi, per garantire un buon risultato, a una risoluzione ottica di 300 dpi; tale valore appare infatti sufficiente a restituire un'immagine digitale ben definita, che potrà egregiamente essere impiegata in sostituzione dell'originale. È chiaro che a tale livello di risoluzione le dimensioni del file aumenteranno in misura considerevole con il crescere delle dimensioni dell'originale; l'economia complessiva del sistema sconsiglierebbe perciò l'acquisizione a valori ottici troppo elevati, in quanto si otterrebbe così un sovradimensionamento degli archivi, e senza un significativo beneficio in termini di rapporto costi/qualità/gestione.

Per tutti quei documenti dei quali si intenda invece distribuire innanzitutto il contenuto – e ciò in modo a volte indipendente da caratteristiche quali pregio e antichità,⁸ e a condizione che la loro leggibilità sia comunque agevole – i valori di risoluzione ottica potranno essere mantenuti entro i 200 dpi, con significativi abbassamenti di tale valore nei casi che i documenti abbiano un formato A4 e che siano leggibili senza alcuna difficoltà. Anche per questo genere di documenti può valere la regola della massima economia del sistema, ossia della ripresa al valore minimo di garanzia.

c) *Archiviazione e supporti*. Se le immagini JPG a bassa risoluzione non sono di difficile archiviazione, in quanto la piccola dimensione dei file fa sì che molte migliaia di immagini possano essere archiviate in un comune disco fisso, altri sono invece i problemi per quel che concerne le immagini TIFF ad alta risoluzione. La digitalizzazione in formato TIFF, a 300 dpi e a 24 bit-colore, di un documento che misuri una cinquantina di centimetri sul lato più lungo può infatti generare un file di circa cinquanta MB; moltiplicando tale misura per le migliaia di documenti che possono essere trattati all'interno di un progetto, si ottengono facilmente archivi dalle dimensioni di centinaia di GB, e la gestione di simili quantità di memoria può presentare non poche difficoltà.

Le soluzioni tecnologiche disponibili in tale prospettiva sono molteplici, ma quelle che presentano il maggior interesse sono le seguenti:

Cd-rom: l'archiviazione su cd-rom presenta i seguenti vantaggi: collaudata affidabilità del supporto, anche in termini di permanenza dei dati; basso costo del processo (acquisizione HW e SW, masterizzazione, duplicazione). Sono limiti oggettivi la scarsa quantità di dati archiviabili su singolo supporto e, di conseguenza, la necessità di organizzare e/o acquisire un sistema, automatico o manuale, di distribuzione degli stessi. Il sistema di distribuzione automatica necessita di una libreria, o juke-box, di cd-rom, che sia gestita da applicativi specifici.

Dvd-rom: può in questo caso risultare meno certa, a causa della scarsa esperienza accumulata in proposito, l'affidabilità del metodo. La maggiore capacità dei dvd rispetto ai cd – circa otto volte tanto – riduce ma non risolve i problemi ➤

⁸ Può essere questo il caso, ad esempio, di antichi testi a stampa soggetti a frequente consultazione dei quali si intenda preservare innanzitutto l'integrità fisica.

connessi alla distribuzione dei dati, e sarà perciò necessario anche in questo caso ricorrere a librerie automatizzate. Il costo di tali librerie è di poco superiore a quello delle stesse in versione cd.

FMD: la sigla identifica una nascente tecnologia che si annuncia, stando alle dichiarazioni dei produttori, quanto mai promettente. Il Fluorescent multy-layer disc ha infatti l'aspetto di un comune cd, salvo che appare trasparente; le informazioni sono infatti archiviate in modo stratificato sullo stesso supporto, così da ottenere una capacità complessiva di 140 GB (ma si parla anche di un prossimo sviluppo fino alla capacità di 1 TB, corrispondente all'incirca al contenuto di 1.500 cd). Altri punti a favore di tale tecnologia sembrano essere l'elevatissima velocità di trasferimento (1 gigabit/sec.), la piena compatibilità verso il basso, le modifiche assai limitate che sarà necessario apportare allo HW attuale. Il prodotto viene annunciato di prossima commercializzazione.

Hard disk di rete: questa tecnologia consiste in uno o più box grandi quanto un comune pc da tavolo, che contengono un hard disk che può arrivare alla capacità di 120 GB (corrispondente a circa 185 cd-rom), una scheda di rete e un lettore di cd-dvd. Il contenuto dei supporti mobili può essere facilmente scaricato nell'hard disk, in modo tale che questo – da solo o in cascata – e i dati in esso contenuti diventino accessibili al pari di qualsiasi altra risorsa di rete. Il vantaggio di tale sistema è l'estrema affidabilità, nonché la velocità di trasferimento dati. A sfavore giocano la limitata (per ora) capacità di memoria e il costo, che a parità di volume dell'archivio risulta più elevato di quello di una libreria di cd-dvd.

4. Immagini digitali: caratteristiche e tipologie

Le immagini digitali si possono definire innanzitutto in base al colore e al numero di bit che "informa" ciascun pixel dell'immagine.

Si hanno così:

Immagini in bianco e nero (1 bit): ciascun pixel contiene un solo bit di informazione, ed è perciò o bianco o nero.

Immagini a 16 colori (4 bit): la combinazione di informazioni può generare 16 toni di grigio o 16 colori.

Immagini a 256 colori (8 bit): la combinazione di informazioni può generare 256 toni di grigio o 256 colori.

Immagini High Color (16 bit): sono formate dalla combinazione dei tre colori fondamentali, il rosso, il verde e il blu, che possono generare da 32.268 a 65.536 colori.

Immagini True Color (24 bit): ciascun pixel contiene una delle 256 tonalità disponibili nei tre canali-colore (rosso, verde, blu), fino a comprendere un totale di 16,8 milioni di colori.

Formato JPG (Joint Photographic Experts Group). Come già accennato in precedenza, il formato JPG non è in alcun

modo adatto alla produzione di copie d'archivio delle immagini digitali, in quanto si tratta di un formato compresso del tipo detto *lossy*, ovvero a perdita di qualità (una perdita che non può essere recuperata); il fattore di compressione può variare, e si possono così ottenere immagini le cui dimensioni sono ridotte di venti e più volte rispetto all'originale. L'immagine JPG sarà *in ogni caso* un'immagine compressa, in quanto la perdita sarà effettiva anche se il coefficiente di compressione – predeterminabile attraverso i menu operativi dei diversi programmi di trattamento immagini – sarà apparentemente impostato uguale a zero. Un'immagine compressa al 10% non sarà perciò diversa, in sostanza, da un'immagine compressa allo 0%. Inoltre, ogni volta che la stessa immagine viene aperta e salvata, si verifica un'ulteriore processo di compressione, il che conduce a lungo andare – in caso di salvataggi ripetuti – a uno scaldamento dell'immagine. Per contro, tale formato risulta quanto mai adatto in tutti quei casi nei quali la risoluzione dell'immagine digitale non sia un fattore determinante, ovvero quando sia sufficiente il suo *riconoscimento* (questo caso si verifica quando si hanno delle immagini JPG che potremmo definire "di primo livello"); una volta riconosciuta l'immagine e verificata in tal modo la sua generica rispondenza ai criteri della ricerca, sarà necessario, per poterla effettivamente esaminare, passare alla sua versione non compressa.

Formato TIFF (Tag Image File Format). È il formato che viene più largamente riconosciuto, in maniera indipendente dalle piattaforme e dai programmi impiegati. Non vi è in questo caso perdita di informazioni, e anche la versione di TIFF (lievemente) compressa detta LZW non comporta perdite. Occorre piuttosto osservare che soltanto le specifiche del formato TIFF 6.0 sono disponibili in forma ufficiale, mentre quelle delle versioni precedenti, TIFF 4.0 e 5.0, che sono ancora di largo utilizzo, sono divulgate in una versione non ufficiale, la quale non offre tutte le garanzie necessarie, soprattutto in vista di futuri passaggi ad altre tecnologie HW e SW.⁹

Formato PNG (Portable Network Graphics). Si tratta di un formato ancora scarsamente diffuso, che presenta tuttavia due notevoli vantaggi: non comporta perdita di informazione e le sue dimensioni sono più piccole del 30% circa rispetto al corrispondente file TIFF LZW. Il suo impiego per la creazione di *master files* è perciò accettabile, anche se alcune piattaforme e alcuni programmi non sono ancora in grado di riconoscerlo.

Per riassumere, e a semplice titolo d'esempio, si possono indicare i seguenti valori di raffronto, riferiti a un'immagine digitale di 1943 x 1702 pixels: TIFF 9.9 MB, TIFF LZW 8.4 MB, PNG 6.5 MB, JPG compresso al 75% 500Kb.¹⁰

5. Il questionario

Lontani dall'immaginare che le note precedenti possano

⁹ Sull'argomento, si vedano le anticipazioni dello studio commissionato alla Cornell University Library dal Council on Library and Information Resources, anticipazioni pubblicate nell'articolo *Risk management of digital information: a file format investigation*, "RLG DigiNews", 4 (2000), 3, alla pagina <<http://www.rlg.org/preserv/diginews/diginews4-3.html>>.

¹⁰ Tali dati sono citati dalla pagina <<http://www.scantips.com/basics09.html>>.

aver esaurito la rassegna dei temi e delle questioni relativi all'argomento, ma ritenendo in ogni caso opportuno proporre strumenti che possano essere operativamente utili, si è pervenuti alla compilazione di un questionario che potrà – ci si augura – servire alla redazione di uno schema progettuale per le attività di digitalizzazione nelle biblioteche. L'impiego di tale questionario potrà rivelarsi efficace per definire il contorno generale del progetto e alcune sue caratteristiche specifiche e, in particolare, per:

1) definire l'ambito generale di sviluppo del progetto in relazione agli scopi e alle risorse disponibili da un lato, alle conoscenze e alle esperienze generalmente acquisite in ambito bibliotecario dall'altro;

2) valutare i fattori tecnici che possono influenzare la scelta dei fondi da sottoporre al trattamento digitale;

3) definire lo schema generale delle risorse umane necessarie per il progetto;

4) valutare la dimensione delle piattaforme tecnologiche da adottare;

5) offrire uno strumento per il confronto operativo fra le istituzioni committenti e i soggetti pubblici e privati chiamati a concorrere alla realizzazione del progetto;

6) fornire un'informazione di base sulle metodologie e le tecnologie da adottarsi nel corso dell'attività;

7) definire alcune questioni tecniche di base relative al formato digitale (in particolare a quello grafico). ■

Ringrazio Stefania Rossi Minutelli per la paziente rilettura e per i preziosi suggerimenti

QUESTIONARIO

1. ANAGRAFE BIBLIOTECA

Nome _____

Natura istituzionale della biblioteca _____

Direttore/Direttrice _____

Indirizzo _____

Telefono / Fax _____

E-mail _____

Indirizzo Home Page _____

Nome del progetto _____

Ufficio al quale fa capo il progetto _____

Responsabile _____

2. SCOPO DELLA DIGITALIZZAZIONE

A quale scopo si intendono impiegare i documenti in formato digitale? (sono possibili risposte multiple)

ACCESSO	<input type="checkbox"/>
ARCHIVIAZIONE	<input type="checkbox"/>
PROTEZIONE DEGLI ORIGINALI	<input type="checkbox"/>
SERVIZI DI DOCUMENT DELIVERY	<input type="checkbox"/>
PUBBLICAZIONE	<input type="checkbox"/>

Quali benefici si intendono ottenere per mezzo della conversione?

_____ ➤

ARGOMENTI

Si prevede un'effettiva domanda di accesso ai documenti in formato digitale?

Quali sono gli scopi principali che il progetto di digitalizzazione intende ottenere?

Si ritiene di maggiore importanza la protezione degli originali o la maggiore facilità di accesso agli stessi che il formato digitale può offrire?

I documenti in formato digitale dovranno sostituire del tutto gli originali, oppure saranno di supporto alla consultazione degli stessi?

I documenti in formato digitale arricchiranno archivi digitali già esistenti, oppure costituiranno lo strumento di accesso privilegiato per la consultazione di determinate tipologie di materiale e/o di specifiche aree tematiche?

3. UTENZA

Uno degli effetti che ogni progetto di digitalizzazione può produrre è l'ampliamento dell'utenza. Si ritiene di poter indicare che un tale ampliamento sia uno degli scopi principali che il progetto si propone?

A QUALE TIPO DI UTENTE SI RIVOLGE INNANZITUTTO IL PROGETTO?

INTERNO (PERSONALE INTERNO E UFFICI)	<input type="checkbox"/>
ESTERNO LOCALE (LETTORI IN SEDE)	<input type="checkbox"/>
ESTERNO REMOTO (OPAC, INTERNET)	<input type="checkbox"/>
UTENTI GENERICI	<input type="checkbox"/>
UTENTI CON INTERESSI E PREPARAZIONE SPECIALISTICI	<input type="checkbox"/>

Esistono già in biblioteca postazioni informatiche per la consultazione diretta e/o assistita da parte dei lettori?

SÌ ☐ NO ☐

4. PERSONALE ADDETTO

Esistono nella biblioteca professionalità in grado di seguire gli sviluppi tecnici del progetto e di coordinare gli interventi?

SÌ ☐ NO ☐ UNITÀ _____

La biblioteca dispone di personale da adibire alla movimentazione e alla preparazione dei materiali da trattare?

SÌ ☐ NO ☐ UNITÀ _____

La biblioteca dispone di personale esperto nell'uso degli strumenti informatici e di quelli fotografici?

SÌ ☐ NO ☐ UNITÀ _____

La biblioteca dispone di personale da adibire a compiti di assistenza per la consultazione dei documenti digitali da parte del pubblico?

SÌ ☐ NO ☐ UNITÀ _____

La biblioteca dispone di personale da adibire alle operazioni di indicizzazione e/o catalogazione dei materiali trattati?

SÌ ☐ NO ☐ UNITÀ _____

5. LOGISTICA

La biblioteca intende realizzare le operazioni di fotografia e/o digitalizzazione in sede oppure fuori sede?

IN SEDE ☐ FUORI SEDE ☐ ENTRAMBE ☐

La biblioteca dispone di un locale adatto nel quale possa essere allestito un laboratorio per le riprese fotografiche e/o il trattamento digitale degli originali?

SÌ ☐ NO ☐

6. CARATTERI INTRINSECI ED ESTRINSECI DEI MATERIALI DA SOTTOPORRE ALLA DIGITALIZZAZIONE

Indicare, fra quelli di seguito elencati, i materiali che dovranno essere trattati; indicare anche la percentuale approssimativa di ciascun tipo e genere in rapporto al totale dei materiali da trattare.

Importante: I materiali sciolti devono essere conteggiati singolarmente, anche se riuniti in cartelle, buste, ecc.; quelli legati in volume devono essere conteggiati a pagina, indipendentemente dal sistema di paginazione originario.

N.B. Per *testo* si intendono quei documenti nei quali l'informazione testuale sia prevalente. Per *immagine* si intendono quei documenti nei quali sia prevalente l'informazione visuale. Per *misto* si intendono quei documenti nei quali l'informazione visuale sia corredata in modo significativo da elementi testuali; possono appartenere a quest'ultima classificazione i materiali cartografici manoscritti o a stampa contenenti numerosi toponimi o altri documenti visuali con significativa presenza di iscrizioni, didascalie, legende, ecc.

6.1 TESTO. SUPPORTI

TESTO MANOSCRITTO SU PERGAMENE SCIOLTE	_____	%	_____
TESTO MANOSCRITTO SU PERGAMENE LEGATE	_____	%	_____
TESTO MANOSCRITTO SU CARTE SCIOLTE	_____	%	_____
TESTO MANOSCRITTO SU CARTE LEGATE	_____	%	_____
TESTO DATILOSCRITTO SU CARTE SCIOLTE	_____	%	_____
TESTO DATILOSCRITTO SU CARTE LEGATE	_____	%	_____
TESTO A STAMPA SU CARTE SCIOLTE	_____	%	_____
TESTO A STAMPA SU CARTE LEGATE	_____	%	_____

6.2 TESTO. DIMENSIONI

TESTO SU FOGLIO DI DIMENSIONI FINO AD A4	_____	%	_____
TESTO SU FOGLIO DI DIMENSIONI FINO AD A3	_____	%	_____
TESTO SU FOGLIO DI DIMENSIONI FINO AD A2	_____	%	_____
TESTO SU FOGLIO DI DIMENSIONI FINO AD A1	_____	%	_____
TESTO SU FOGLIO DI DIMENSIONI SUPERIORE AD A1	_____	%	_____

6.3 IMMAGINI. GENERE

DISEGNO (GRAFITE, PENNA, ECC.)	_____	%	_____ ➤
--------------------------------	-------	---	---------

DISEGNO A COLORI	_____	%	_____
GRAFICA AL TRATTO	_____	%	_____
GRAFICA A TONI DI GRIGIO	_____	%	_____
GRAFICA A COLORI	_____	%	_____
MINIATURE	_____	%	_____
FOTOGRAFIE IN BIANCO E NERO	_____	%	_____
FOTOGRAFIE A COLORI	_____	%	_____

6.4 IMMAGINI. DIMENSIONI

IMMAGINI SU FOGLIO DI DIMENSIONI FINO AD A4	_____	%	_____
IMMAGINI SU FOGLIO DI DIMENSIONI FINO AD A3	_____	%	_____
IMMAGINI SU FOGLIO DI DIMENSIONI FINO AD A2	_____	%	_____
IMMAGINI SU FOGLIO DI DIMENSIONI FINO AD A1	_____	%	_____
IMMAGINI SU FOGLIO DI DIMENSIONI SUPERIORE AD A1	_____	%	_____

6.5 MISTO. GENERE

MISTO MANOSCRITTO A PENNA	_____	%	_____
MISTO MANOSCRITTO A COLORI	_____	%	_____
MISTO A STAMPA AL TRATTO	_____	%	_____
MISTO A STAMPA A TONI DI GRIGIO	_____	%	_____
MISTO A STAMPA COLORATO	_____	%	_____

6.6 MISTO. DIMENSIONI

MISTI SU FOGLIO DI DIMENSIONI FINO AD A4	_____	%	_____
MISTI SU FOGLIO DI DIMENSIONI FINO AD A3	_____	%	_____
MISTI SU FOGLIO DI DIMENSIONI FINO AD A2	_____	%	_____
MISTI SU FOGLIO DI DIMENSIONI FINO AD A1	_____	%	_____
MISTI SU FOGLIO DI DIMENSIONI SUPERIORE AD A1	_____	%	_____

TOTALE TESTO

TOTALE IMMAGINI

TOTALE MISTI

_____	%	_____
_____	%	_____
_____	%	_____

7. CONSERVAZIONE

Indicare quali problemi di conservazione caratterizzino determinati tipi di materiali, specificando anche quali siano i materiali eventualmente affetti da particolari problemi di conservazione e la loro quantità espressa, se possibile, in percentuale

Fragilità del supporto (materiale di pregio antico e moderno) % _____
Descrizione

Fragilità del supporto (materiale non di pregio: ad es. giornali) % _____
Descrizione

Documenti di grandi dimensioni conservati ripiegati % _____
Descrizione

Volumi con legatura molto stretta al dorso
Descrizione

% _____

Volumi con legatura fragile, che non sopportano l'apertura a 180°
Descrizione

% _____

Documenti di difficile lettura per sbiadimento, dilavazione,
ecc. delle informazioni testuali o grafiche contenute
Descrizione

% _____

Documenti di difficile lettura per forte scurimento della carta dovuto
a fattori chimici e/o ambientali (acidità, macchie di umidità, *foxing*)
Descrizione

% _____

Documenti di difficile lettura per acidità dell'inchiostro, con trasparenza
dello stesso ed eventuale foratura della carta
Descrizione

% _____

8. DOTAZIONI HARDWARE, SOFTWARE E DI RETE

Indicare le dotazioni già esistenti destinate ad essere impiegate nel progetto

COMPUTER	UNIX	QUANTITÀ _____	TIPO _____
	PC	QUANTITÀ _____	TIPO _____
	MAC	QUANTITÀ _____	TIPO _____
	ALTRO	QUANTITÀ _____	TIPO _____
MONITORS 17"		QUANTITÀ _____	
MONITORS 20" O SUPERIORE		QUANTITÀ _____	
SCHUDE GRAFICHE CON ALMENO 8 MB DI VRAM		QUANTITÀ _____	
LETTORI CD-ROM		QUANTITÀ _____	VELOCITÀ _____
LETTORI DVD		QUANTITÀ _____	VELOCITÀ _____
LIBRERIE CD/DVD		QUANTITÀ _____	

ARGOMENTI

MASTERIZZATORI CD	QUANTITÀ _____	VELOCITÀ _____
MASTERIZZATORI DVD	QUANTITÀ _____	VELOCITÀ _____
ALTRI SUPPORTI DI ARCHIVIAZIONE DATI		
DAT	_____	
JAZ 1/2GB	_____	
ALTRO	_____	
STAMPANTI		
JET BN	QUANTITÀ _____	FORMATO (A4/A3) _____
JET COLORE	QUANTITÀ _____	FORMATO (A4/A3) _____
LASER BN	QUANTITÀ _____	FORMATO (A4/A3) _____
LASER COLORE	QUANTITÀ _____	FORMATO (A4/A3) _____
CONNESSIONE IN RETE LOCALE (LAN)	_____	
CONNESSIONE INTERNET	_____	
SOFTWARE		
SISTEMA OPERATIVO UNIX	VERSIONE _____	
PC WINDOWS	VERSIONE _____	
MAC OS	VERSIONE _____	
ALTRO SOFTWARE	_____	
VIEWER	PRODOTTO _____	
TRATTAMENTO IMMAGINI	PRODOTTO _____	

9. INDICIZZAZIONE, CATALOGAZIONE

Quali strumenti di ripertimento/indicizzazione/catalogazione sono già disponibili per i materiali da trattare?

CATALOGO CARTACEO	_____
DATABASE BIBLIOGRAFICO	_____
INDICE CARTACEO	_____
INDICE INFORMATIZZATO	_____
INVENTARIO CARTACEO	_____
INVENTARIO INFORMATIZZATO	_____
NESSUNO	_____

Nel caso in cui non esista nessuno strumento di reperimento/indicizzazione/catalogazione si prevede che tali strumenti possano essere predisposti in vista del progetto?

SÌ ☐ NO ☐

La predisposizione di tali strumenti sarà finalizzata dal progetto o sarà curata dalla biblioteca quale attività ordinaria?

Il trattamento digitale dei materiali produrrà un numero *n* di files raggruppati in directories e subdirectories che ricostruiranno la sequenza fisica degli originali. Si ritiene che tale livello di sviluppo sia sufficiente per gli scopi del progetto?

SÌ ☐ NO ☐

Il progetto prevede invece che i materiali tratti possano confluire in un sistema integrato con funzioni di inserimento dati (pre-catalogazione, catalogazione, ecc.), interrogazioni degli stessi, visualizzazione e stampa delle immagini?

SÌ ☐ NO ☐

Un simile sistema è già disponibile in biblioteca?

SÌ ☐ NO ☐

se sì, specificare

Quale programma di catalogazione si intende impiegare per il progetto?

Le eventuali operazioni di catalogazione saranno affidate a personale interno oppure esterno?

Tutti i documenti sottoposti a digitalizzazione dovranno essere catalogati (ad es. nel caso di spogli) oppure la catalogazione dovrà essere effettuata soltanto per il livello bibliografico principale?

È possibile quantificare separatamente il numero di documenti oggetto del trattamento digitale e della catalogazione?

TRATTAMENTO DIGITALE _____
CATALOGAZIONE _____

Si prevede che i documenti sottoposti a trattamento digitale siano catalogati

PRIMA DELLA DIGITALIZZAZIONE _____
CONTESTUALMENTE ALLA DIGITALIZZAZIONE _____
DOPO LA DIGITALIZZAZIONE _____

Il finanziamento del progetto include le operazioni di catalogazione?

SÌ ☐

NO ☐